Journal of Educational Psychology
1985, Vol. 77, No. 5, 562–571

0022-0663/85/$00.75

Postprimary Education Has Little Impact on Informal Reasoning

D. N. Perkins
Harvard University

The present study examined whether postprimary education enhances informal reasoning skills, operationalized as skill in the construction of arguments about everyday issues. Eight groups of 40 subjects, balanced for sex, ranged over high school, college, graduate school, and nonstudents with and without a bachelor's degree. Each subject gave oral arguments on two issues; responses were scored for overall quality, number of lines of argument, and several other factors. Analysis disclosed a borderline statistically significant impact of high school, college, and graduate school. However, both level of performance and rates of gain with education were much lower than one would hope. It is argued that present educational practices do little to foster the development of informal reasoning skills; education redesigned for this purpose could have a much greater impact.

Schooling traditionally aims to prepare students for life beyond academe. To this end, schools seek to equip students in several particular areas of knowledge and skill—reading, mathematics, and history for example. However, the aspirations of education, particularly after the primary grades, go beyond this. One hopes that students will emerge from 12 or more years of study not just better able to read, write, reckon, or recall particular facts, but to *think*.

Does education succeed in this higher mission? This article reports the results of an investigation of the impact of high school, college, and graduate school on a certain kind of thinking—everyday or informal reasoning. Informal reasoning, as defined in this line of inquiry (see also Perkins, 1985; Perkins, Allen, & Hafner, 1983), involves considering a claim and seeking reasons with a nonformal bearing on the claim, pro or con, in an attempt to resolve the truth of the claim. Informal reasoning stands in contrast to formal reasoning, characteristic of mathematics, syllogistic reasoning, and probabilistic reasoning, in which the conclusion follows by strict deduction or calculation from given premises. In informal reasoning, reasons typically occur on both sides of the case, no one line of argument settles the truth of the claim, and no computational procedure assigns the claim a numerical probability. The reasoner must weigh and synthesize to best judge the soundness of the claim.

Most reasoning that people do in everyday and even academic life is informal. Decision-making situations from purchasing a car to resolving which experimental design to use typically require people to reason out the pros and cons of the options. Scholarly pursuits in the humanities call for advancing a thesis and arguing its merits on nonformal grounds. To be sure, in mathematics and the mathematical side of the sciences, formal arguments play an important role, but informal argument figures as well: The elegance of a theory, the appropriateness of its axioms, its range of application, and how well it does compared with rival theories are all matters that characteristically involve considerable informal argumentation. We have no syllogisms for settling such disputes.

Psychological research has paid most heed to the nature and development of formal reasoning skills, for instance syllogistic reasoning (e.g., Falmagne, 1975; Johnson-

This research was supported by the Spencer Foundation and National Institute of Education Grant NIE-G-83-0028, *Learning to Reason*. The views expressed in this article do not necessarily reflect the position or policy of the supporting agencies.

I thank Richard Allen and James Hafner for their extensive contribution to all phases of the work reported in this article.

Requests for reprints should be sent to D. N. Perkins, Graduate School of Education, 315 Longfellow Hall, Harvard University, Cambridge, Massachusetts 02138.

Laird, 1983; Revlin & Mayer, 1978; Wason & Johnson-Laird, 1972), probabilistic reasoning (e.g., Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980; Slovic, Fischhoff, & Lichtenstein, 1977), reasoning in mathematics and physics problems (e.g., Chi, Feltovich, & Glaser, 1981; Greeno, 1983; Larkin, 1983; Larkin, McDermott, Simon, & Simon, 1980; Newell & Simon, 1972; Schoenfeld, 1980, 1982; Schoenfeld & Herrmann, 1982), and formal reasoning in the Piagetian sense (Inhelder & Piaget, 1958). The emphasis on formal reasoning seems to have been motivated by the manifest difficulties people have with it, the investigative convenience afforded by a formal criterion of correct inference, and the relative ease of implementing models of formal reasoning processes on computer or at least conceptualizing formal reasoning in information-processing terms.

Whatever the factors involved, the neglect of informal reasoning is unfortunate. As emphasized already, most of the reasoning people do has an informal character. Moreover, if people are not very good at formal reasoning, they are perhaps not very good at informal reasoning either. For instance, research has shown that people tend to overweight the influence of salient individuals on a situation; increase their commitment to their original positions in response to mixed evidence, when it should lead them to reduce their confidence; and preserve a belief even after the evidence on the basis of which it was formed has been thoroughly discredited (Ross & Anderson, 1982). Some might argue that to study formal reasoning is to study informal reasoning, because the latter simply uses formal mechanisms more loosely. However, Perkins (1985) and Perkins et al. (1983) have argued at length that informal reasoning calls for generating and weighing lines of argument in ways not required by formal argument.

Figuring as it does in so many aspects of academic and nonacademic life, informal reasoning skill clearly presents a natural and important educational objective. The current resurgence of interest in critical thinking reflects widespread recognition of this point. One might hope that education, particularly at the college and graduate school level, has some impact on informal reasoning abilities. After all, students receive considerable exposure to arguments about issues and gain some experience in constructing such arguments themselves, through essay assignments. At the same time, one may doubt whether conventional education provides sufficient focus and practice to enhance informal reasoning very much. Moreover, factors concerning the nature of expertise and intelligence may limit the improvement of informal reasoning skills, a point examined at greater length in the discussion section of this paper. With these uncertainties in mind, a study was designed to appraise directly the impact of education on informal reasoning.

Such an inquiry required the design of a task to operationalize informal reasoning and of measures to gauge quality of reasoning. The chosen task asked subjects to consider public issues not demanding extensive knowledge and to develop a position and supporting arguments on them. Simple measures of argument quality were devised to appraise the subjects' performances. Groups of subjects were drawn from the first and fourth years of high school, college, and graduate school, as well as from individuals who had finished their schooling, to provide a sample ranging across and beyond postprimary education. The same data also were used to analyze the nature of difficulties in informal reasoning and their relation to difficulties in formal reasoning (Perkins et al., 1983), but the present study addressed educational impact only.

Method

Subjects

There were 320 subjects divided into 8 groups of 40 subjects each, with each group balanced for sex. The groups were as follows: first year high school students, fourth-year high school students, first-year college students, fourth-year college students, first-year graduate students enrolled in doctoral programs, fourth-year graduate students, nonstudents who had been neither students nor teachers for more than 5 years and who had a high school diploma but not a bachelor's degree, and nonstudents fulfilling the same requirement but with a bachelor's degree. The high school students participated voluntarily; the rest received a moderate fee. Each student group included subjects from at least two different schools. Schools

of exceptional reputation were avoided so that the results would better reflect normal education, except that around half of the graduate students attended Harvard University. The nonstudent subjects all came from a middle-class suburb of Boston.

Procedure

Each subject participated in an interview conducted by one of the investigators. The interview lasted around an hour and a half. First the investigator recorded such basic information as age, years of education, sex, and academic major if any. Then the investigator presented the subject with an issue typed on a sheet of paper. The investigator asked the subject to think about this issue alone for 5 min, reaching a conclusion if possible. The investigator provided scratch paper, encouraging the subject to make any notes that would be useful. After the 5 min, the subject explained any conclusion and the argument for it. Whenever the subject seemed to be finished, the experimenter encouraged the subject to say more, to ensure that the subject gave as full an account as possible. The subject's remarks were tape recorded.

After that, the investigator posed a number of follow-up questions designed to further probe the subject's reasoning. The investigator asked the subject to indicate how much thought prior to the experiment the subject had given to the issue, eliciting a three-way distinction that became the prior-thought variable: (a) no prior thought, (b) some but less than during the 5 min, and (c) at least as much as the 5 min. This variable was employed during the analysis to check for the influence of previous thinking on performance. The experimenter also asked whether the subject found the 5 min sufficient to think about the issue, given that no additional information was available. The other follow-up questions included queries about how certain the subject felt of his or her conclusion and how the subject explained the connection between one reason the subject had given and his or her conclusion.

The experimenter repeated this entire cycle again, introducing a new issue, providing the 5 min, collecting the argument and pursuing the follow-up questions. Finally, the investigator administered the Slosson Intelligence Test, a short-form IQ test (Slosson, 1981).

The issues used in the research were chosen for being genuinely vexed issues with some currency at the time the data were gathered. Four issues were employed in a counterbalanced design. The issues were selected from a larger pool, after piloting, because they permitted elaborate arguments on both sides of the case, led to divided opinions, proved accessible even to the first-year high school group, and did not depend for their analysis on background knowledge that varied greatly across the subject population. Briefly stated, the issues were as follows:

1. Would restoring the military draft significantly increase America's ability to influence world events?
2. Does violence on television significantly increase the likelihood of violence in real life?
3. Would a proposed law in Massachusetts requiring a five cent deposit on bottles and cans significantly reduce litter?
4. Is a controversial modern sculpture, the stack of bricks in the Tate Gallery, London, really a work of art?

Scoring

The tape-recorded responses of the subjects were scored on several scales that provided measures of the quality of the subjects' arguments. All scoring was performed by two judges independently. After the scoring was completed, each scale was examined for the correlation between the judges' scores and the correlation between subjects' performance on the first and the second issue posed. The latter examined whether the scales measured a property of the subject or merely a property of the individual performance.

A couple of scales were discarded for poor interjudge agreement and one for a very low first issue–second issue correlation. An additional scale was set aside as redundant. Six scales remained, on which the subsequent analysis focused. The six scales were as follows:

Sentences. The number of sentences in the subject's argument was counted as a simple measure of elaboration. An alternative measure, number of premises, was set aside because it correlated highly with number of sentences but had a slightly lower correlation between judges. Another potential misgiving about the measure was that some subjects might have padded their responses with irrelevancies more so than others. To check against this possibility, the judges also provided an irrelevant-sentence count. However, the irrelevant sentences measure both correlated highly with the sentences measure and, unlike the sentences measure, showed poor first isssue–second issue correlation and poor correlation with rating. Redundant and of questionable validity, the irrelevant sentences measure was dropped; the sentences measure was used unmodified for the data analysis.

Lines of argument. Each judge also counted the lines of argument in each subject's response to an issue. Lines of argument were distinctly different ways of arguing the point in question. For example, suppose a subject argued in favor of a deposit law reducing litter both because people would return their bottles for the 5 cents and because such a law supposedly had reduced litter in Vermont. The subject would receive credit for two lines of argument. If the subject elaborated each of these extensively, he or she received a higher sentence count, but still only credit for two lines of argument. Thus lines of argument and sentences served as measures of breadth of search and extent of search, in rough analogy to measures of flexibility and fluency (Guilford & Hoepfner, 1971).

Objections. The judges counted how many objections a subject raised to his or her own position in an argument. Each objection contributed one point, regardless of whether the subject merely mentioned it or offered a rebuttal. Objections provided a measure of the extent to which subjects considered the other side of the case. It should be recalled that the issues were chosen to be vexed and pretested to ensure that elaborate arguments existed on both sides of the case. Accordingly, a low objection score indicated failure to discern or express the arguments on the other side, not the absence of arguments on the other side.

Prompts. Many subjects tended to drift away from

the issue under discussion as they offered their arguments. The procedure allowed the experimenter to prompt a subject to return to the given issue. The prompts measure counted the number of times for a given argument the experimenter took this option. Thus, this scale provided an estimate of the subject's ability and/or willingness to remain focused on the issue.

Explanation. In following up on a subject's argument, the experimenter singled out one reason the subject gave and asked the subject to explain how the reason supported the conclusion. The reason was selected by the experimenter according to a priority list of the reasons subjects typically offered, based on pilot work with the issues. Both judges scored on a 5-point scale the adequacy of a subject's explanation. This provided an indication of the subject's ability to explicate the logic of his or her argument.

Rating. Each judge, upon listening to a tape-recorded argument, rated it on a 5-point scale for overall quality. This judgment was made prior to writing down scores for the other measures, so that they would be less likely to influence the overall quality rating directly. The rating provided a way to check whether the other measures captured at least in part what one means intuitively by a good argument.

Results

Design Validity

A study such as the present one that seeks a sample of fairly "natural" performances is subject to many hazards, not all of which can be eliminated entirely by niceties of design. However, some partial checks against certain of these hazards were possible:

Matching of first- and fourth-year samples. A spurious impact of education on informal reasoning might appear if, for example, a heavy rate of dropout in high school, college, and graduate school led to the fourth-year samples including better intellectual performers in general. In such a case, what in fact was a selection effect would look like an educational effect. The IQ measure provided a sample-matching criterion that made this unlikely. There was no significant difference between the mean IQs of the first- and fourth-year samples at the high school, college, or graduate school levels.

Sufficient time to think. Because vexed issues were chosen, it is reasonable to wonder whether subjects had time to explore those issues and develop arguments. Perhaps time would impose a ceiling on the best subjects. However, this did not appear to be a serious concern. First of all, subjects were asked whether they had enough time in the 5 min provided to think through the issue, given that no more information was available. Over all groups, 72% felt that the 5 min sufficed. Although one might argue that in principle the issues could be explored at length, in fact, most subjects ran out of ideas rather quickly. In addition, it must be remembered that the 5 min were not really the sole opportunity for subjects to explore the issue. Subjects often extended their arguments while reporting them and, as noted earlier, they were encouraged to continue developing their arguments as long as they could. Finally, sufficient time or not, the methodology disclosed intergroup contrasts in reasoning performance, as discussed below.

Validity of the measures. Three correlations were calculated to examine the validity of the measures: between the two judges, between the subjects' scores on the first and second issue they addressed, and between the overall rating measure and the other five measures. As noted earlier, some measures were discarded for a poor showing on one or another of these correlations. The interjudge correlations for the six measures ranged from .57 to .94, all significant at the .001 level or better. The correlations between the first and second issues ranged from .22 to .53, also all significant at the .001 level or better. The lower magnitude of these correlations in comparison with the interjudge correlations was understandable in light of each subject's addressing two different issues that would call upon his or her background knowledge and reasoning strategies in somewhat different ways. Finally, the correlations between rating and the other five measures ranged in absolute value between .33 and .64, with the rating–prompts correlation having an appropriately negative sign. Again, all were significant at the .001 level or better.

Impact of Education

Detecting impact of education called for a comparison between the first- and fourth-

Table 1
First-Year Scores and Gains for High School, College, and Graduate School

	High School			College			Graduate school		
Measure	Yr 1	Change	*p*	Yr 1	Change	*p*	Yr 1	Change	*p*
Sentences	10	4	.025	18	3	*ns*	26	3	*ns*
Lines of argument	1.8	.4	.025	2.9	.1	*ns*	3.3	.3	*ns*
Objections	.6	.2	*ns*	1.1	.1	*ns*	1.3	.6	.025
Prompts	3.3	−1.1	.001	.5	0	*ns*	.2	.1	.025
Explanation	1.8	.5	.05	2.4	.4	.05	3.0	−.1	*ns*
Rating	1.6	.5	.001	2.8	0	*ns*	3.1	.2	*ns*

year students at the high school, college, and graduate school levels. To be sure, graduate students in general score higher than college students, who score higher than high school students. But such differences confound effects of education with selective admission procedures: Only the more intellectually able enter college, and only the especially able enter graduate school. Of course, even contrasting first- with four-year students still confounds education with maturation effects, a point to be discussed later.

Table 1 displays the first-year means for high school, college, and graduate school for each of the six measures, along with the gain from first to fourth year in each case and the level of statistical significance of the gain, based on one-tailed *t* tests. The table discloses a pattern of borderline significant gains. In general, statistically significant gains appear for five out of the six measures in high school, for only one in college, and for two in graduate school.

Students might differ in their reasoning performance for various reasons, for instance general intelligence, prior consideration of the issues posed, impact of education on informal reasoning abilities, or simply maturational factors. To explore such contributing factors, a multiple linear regression was performed on the pooled student data for each of the six measures. The regression variables were IQ, prior thought, years of education, and age. As expected, these variables were substantially intercorrelated, but at least one could determine which were dominant. The results appear in Table 2. IQ proved to be the most influential variable, significant at the .001 level for all six, with standardized coefficients ranging from .32 to .48. Age showed no significant impact. Prior thought proved significant only for the prompts measure, where, understandably, it correlated negatively with need for prompts. Years of education emerged as borderline significant, with significance levels ranging from .01 to around .1 except for the explanation measure, which stood at .4.

In summary, consistent with the *t* tests reported above, the regression disclosed a borderline significant influence of education on informal reasoning ability. The fact that age and prior thought did not in general reach significance suggests that the accumulation of knowledge and general maturation may contribute less than the impact of schooling on general reasoning ability.

The nonstudent data also allowed examining this question, in that years of education was obtained from each nonstudent subject. A multiple linear regression just like that performed for the student subjects was performed for the nonstudents; the results appear in the rightmost section of Table 2. IQ again usually proved to be the most influential variable, although not as influential as among the pooled student groups with the wide range of IQ from high school to graduate students. Years of education had a strong influence only on rating. Age and prior thought never attained a .05 level of statistical significance, although age fell below .1 for two of the six measures. In summary, the regression analysis of the nonstudent data revealed a

Table 2
Regression Analyses of Pooled Student and Pooled Nonstudent Data

Measure	Students pooled: IQ	Students pooled: Age	Students pooled: Years of education	Students pooled: Prior thought	Nonstudents pooled: IQ	Nonstudents pooled: Age	Nonstudents pooled: Years of education	Nonstudents pooled: Prior thought
Sentences								
Standardized coefficient	.36	0.05	.20	.07	.35	.22	.01	−.05
Significance	<.001	*ns*	.10	*ns*	.03	.07	*ns*	*ns*
Lines of argument								
Standardized coefficient	.47	−.01	.22	−.03	.29	.02	.23	−.07
Significance	<.001	*ns*	.06	*ns*	.06	*ns*	*ns*	*ns*
Objections								
Standardized coefficient	.32	−.10	.24	−.06	.10	.03	.17	−.02
Significance	<.001	*ns*	.07	*ns*	*ns*	*ns*	*ns*	*ns*
Prompts								
Standardized coefficient	−.34	.18	−.20	−.18	−.07	.09	−.19	−.02
Significance	<.001	*ns*	*ns*	.01	*ns*	*ns*	*ns*	*ns*
Explanation								
Standardized coefficient	.42	−.07	.12	.03	.57	−.09	.00	.02
Significance	<.001	*ns*	*ns*	*ns*	<.001	*ns*	*ns*	*ns*
Rating								
Standardized coefficient	.48	−.08	.27	.07	.29	−.12	.46	−.01
Significance	<.001	*ns*	.01	*ns*	.02	*ns*	<.001	*ns*

pattern of dominance among the variables similar to but not as sharp as the regression analysis of the student data.

Discussion

Are the levels of performance and gains disclosed by the present analysis satisfactory? If not, can education do better? First consider level of performance. Broadly speaking, the numbers offer no reason to view students as especially competent at informal reasoning. The issues, chosen for a multiplicity of arguments on both sides, received insufficient exploration. The beginning graduate students averaged just 3.3 lines of argument per issue, and the high school freshmen managed only 1.8. The shortfall appears particularly clearly in the objections measure, in which high school freshmen offered .6 objections, and first-year graduate students mustered only 1.3.

To be sure, the unimpressive performance even by graduate students may in part reflect the demand characteristics of the task. It is reasonable to suppose that people would deal at least somewhat more thoroughly with major medical, marriage, or career decisions than with issues of little personal relevance provided out of the blue in the context of an experiment. Nonetheless, everyday experience suggests that people frequently do not reason about major decisions as thoroughly as they might. Moreover, there are many occasions for good reasoning that seem no more motivating than the experimental setting—in which, after all, pains were taken to encourage the subjects to reason well—occasions such as voting on a referendum not central to one's interests, making a purchase decision about an item of moderate cost, writing a term paper, or planning one's activities for the following week.

The pattern of underexploration of an issue finds both parallels and possible explanations in other work. Research on generating plans and hypotheses conducted by Gettys and Engelmann disclosed that subjects typically fall far short in their efforts to explore hypotheses thoroughly to explain a situation or explore plans of action to take in a given situation (Gettys & Englemann, 1983; Gettys, 1983). One reason may be that subjects substantially overestimated the extent to which they had exhausted an issue, which might have led

them to stop prematurely. In the context of studying writing, Bereiter and Scardamalia (1985) addressed the problem of "inert knowledge": In addressing a topic, students access only a fraction of the knowledge they have that bears on the topic.

Regarding the present line of research, Perkins et al. (1983) argued that many reasoners could be characterized as "makes sense epistemologists." Such reasoners proceed to analyze a situation only to the point where the analysis makes superficial sense. For instance, consider whether a 5 cent deposit on bottles and cans would reduce litter. Many younger subjects argued, "Yes, because people would return the bottles and cans for the 5 cents," or "No, because five cents nowadays isn't enough," feeling that these simple scenarios adequately captured the circumstances. To generalize, once the reasoner has evolved a simple mental model with no ostensible flaws, he or she is not likely to critique the model deliberately or consider alternative models. It is as though the reasoning process was driven primarily by an effort to minimize cognitive load and cognitive dissonance rather than by epistemic criteria.

Fortunately, there is evidence that people can learn to do better. For example, Gettys (1983) reported that teaching subjects to analyze a situation and view it more broadly produced gains in the extent to which the subjects explored issues thoroughly. Bereiter and Scardamalia (1985) reported an experiment carried out by their colleague Valerie Anderson in which children, asked to list words they might use in a composition before they started to write, doubled the length of their compositions with a few hours of practice. Apparently the strategy helped them to activate their inert knowledge. The problem of shortfall relates straightforwardly to work on ideational fluency and its enhancement, and a number of efforts to improve ideational fluency on certain sorts of tasks have yielded positive results (see review by Torrance, 1972), although it may be questioned whether this constitutes improvement in overall creativity, the frequent intent of such instruction (Perkins, 1981, Chapter 7).

Now consider the rates of gain in reasoning skills indicated by the present study. Again, one finds little reason to be satisfied. The gains per 3 years of education for high school, college, and graduate school appear in Table 1. Concerning lines of argument, the greatest 3-year gain, .4, occurred in high school, amounting to a little more than one tenth of a line of argument per year of education. The greatest gain for sentences, 4, also occurred in high school; again, the improvement of a little more than one sentence per year of education seems miniscule. The objections measure advanced most in graduate school, by .6 objections over the 3 years. The addition of two tenths of a contrary reason per year of graduate school hardly suggests a substantive gain in critical thinking. In general, the borderline statistically significant gains in reasoning ability documented in this study should not mislead; they do not represent a substantial rate of gain per year of education at any level.

What explains the slow rate of gain? One possible interpretation invokes contemporary research on expertise, which argues that expert performance in any domain depends on a specialized repertoire of knowledge and know-how (e.g., Chi et al., 1981; Larkin et al., 1980; Schoenfeld & Herrmann, 1982; Simon & Chase, 1973). Indeed, a strong expert–novice contrast in the construction of arguments has been documented, with the field of expertise being the Soviet Union (Voss, Tyler, & Yengo, 1983). Presumably, students do improve at reasoning in their specialties as they advance through high school, college, and graduate school, because they acquire the context-specific knowledge underlying expertise. But there is no reason why students should substantially improve in reasoning about general questions such as were chosen for this research, questions that fall outside their expertise.

The relevance of this explanation relates to the distinction between necessary and sufficient conditions. Almost certainly, specific training in the sorts of issues recruited for the present study would yield substantially improved performance: Providing the context-specific knowledge and

know-how characteristic of expertise is sufficient for improving performance and making the learner function more like an expert, as has been demonstrated, for instance, by Schoenfeld and Herrmann (1982) for mathematical problem solving. However, such context-specific expertise arguably is not a necessary condition for improved performance on issues that rely principally on general knowledge of the world.

First of all, as discussed above, people typically underexplore issues, and instruction in general strategies of more thorough exploration helps them to do better. Second, Gettys and Engelmann (1983) offered specific evidence that expertise is not the whole story. They reported a study contrasting graduate students with undergraduate students in generating plans of attack on an open-ended problem in which the graduate students could be considered experts and on one in which they could not. The graduate students explored both problems considerably more thoroughly than did the undergraduate students. Gettys and Engelmann concluded that the graduate students' skill depended in considerable part on general divergent thinking abilities and strategies rather than on content-specific expertise.

The latter point suggests yet another reason why informal reasoning skills might improve but slowly during the postprimary years of schooling: Divergent thinking abilities are often considered an aspect or manifestation of intelligence (e.g., Guilford & Hoepfner, 1971), and intelligence, according to some views at least, lies close to the "hardware" level of the human organism and does not lend itself to substantial change. Such a position gains credence from research that suggests a pervasive *g* factor in human intellectual performance and relates *g* to relatively simple and global parameters of the human information processing system, such as those associated with reaction time (Jensen, 1984).

Here again, the distinction between sufficiency and necessity applies. Without entering the debate on what intelligence is and how subject it might be to change (cf. Detterman & Sternberg, 1982; Whimbey, 1975), one can certainly acknowledge that a contrast in intelligence, as normally measured, typically suffices for better informal reasoning performance. Indeed, the present evidence is strongly consistent with this point: In the regression analyses, IQ proved to be the most influential variable, with standardized coefficients of regression up to .48. However, this does not mean that general intelligence must be improved in some near-"hardware" sense to enhance informal reasoning. One might improve informal reasoning technique, and hence performance, without affecting intelligence at a fundamental level.

Because both increased intelligence and enhanced context-specific expertise are sufficient but arguably not necessary for better informal reasoning, there is no reason to regard the weak gains with education reported here as inevitable. Rather than looking to limitations of human cognitive functioning, one might seek an explanation in shortcomings of the educational process. Broadly speaking, most educational practice does little to prepare students for reasoning out open-ended issues. Much of education does not deal at all with the critical consideration of issues. To be sure, at the university level, critical examination of issues becomes more important—but who does the critical examining? The professor may give his or her critical overview; the text and other readings may expose students to a diversity of arguments on an issue. But neither one of these constitutes direct practice in generating lines of argument, examining both sides of the case, or elaborating and testing out particular lines of argument against one's general knowledge.

The essay assignment is perhaps the only frequently assigned task in which students might practice for themselves such investigative thinking. However, several limitations are immediately apparent. Most courses call for an essay only once a term. Many students meet the demand by papers that summarize and, perhaps, synthesize, without really developing an argument. Professors in particular subject areas rarely provide explicit guidance in how to develop and argue a viewpoint. Instruction in writ-

ing, quite common in college settings, might do so, but students exhibit many writing difficulties besides those directly involved with crafting an argument, and writing instructors naturally tend to treat the range at the expense of any one.

As to graduate education, where the emphasis presumably falls on developing an independent, professionally able thinker, at least two factors seem likely to limit progress. First, to some extent, the circumstances mentioned above apply to graduate as well as undergraduate study. Second, graduate education aims to create experts. But, as already discussed, expertise tends to be context-bound. Whatever reasoning skills a graduate student acquires are likely to be well tuned to the professional discipline and exercised only in that context.

In summary, to whatever limits that intelligence in a near-"hardware" sense and the nature of expertise may impose on the development of informal reasoning skills, one certainly has to add another limiting factor: the lack of exercise and explicit instruction in current educational practice. If the conduct of education routinely emphasized general skills of informal reasoning throughout the subject matters, a substantially greater rate of improvement in informal reasoning skills might appear at all levels. General intelligence and the context-bound nature of expertise presumably put a ceiling on what gains might accrue. But there is no reason to believe that present instructional practices lift students close to that ceiling.

References

Bereiter, C., & Scardamalia, M. (1985). Cognitive coping strategies and the problem of inert knowledge. In S. S. Chipman, J. W. Segal, & R. Glazer (Eds.), *Thinking and learning skills: Vol. 2. Current research and open questions.* Hillsdale, NJ: Erlbaum.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Detterman, D. I., & Sternberg, R. J. (Eds.). (1982). *How and how much can intelligence be increased.* Norwood, NJ: Ablex.

Falmagne, R. J. (Ed.). (1975). *Reasoning: Representation and process in children and adults.* Hillsdale, NJ: Erlbaum.

Gettys, C. F. (1983). *Research and theory on predecision processes.* Norman, OK: Decision Processes Laboratory, University of Oklahoma.

Gettys, C. F., & Engelmann, P. D. (1983). *Ability and expertise in act generation.* Norman, OK: Decision Processes Laboratory, University of Oklahoma.

Greeno, J. G. (1983). Conceptual entities. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 227–252). Hillsdale, NJ: Erlbaum.

Guilford, J. P., & Hoepfner, R. (1971). *The analysis of intelligence.* New York: McGraw-Hill.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking.* New York: Basic Books.

Jensen, A. R. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures, 7,* 93–118.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Larkin, J. H. (1983). The role of problem representation in physics. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 75–98). Hillsdale, NJ: Erlbaum.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Modes of competence in solving physics problems. *Cognitive Science, 4,* 317–345.

Newell, A., & Simon, H. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Perkins, D. N. (1981). *The mind's best work.* Cambridge, MA: Harvard University Press.

Perkins, D. N. (1985). Reasoning as imagination. *Interface, 16,* 14–26.

Perkins, D. N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), *Thinking: The expanding frontier* (pp. 177–189). Philadelphia, PA: Franklin Institute Press.

Revlin, R., & Mayer, R. E. (Eds.). (1978). *Human reasoning.* Washington, DC: V. H. Winston & Sons.

Ross, L., & Anderson, C. A. (1982). Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.)., *Judgment under uncertainty: Heuristics and biases* (pp. 129–152). Cambridge, England: Cambridge University Press.

Schoenfeld, A. H. (1980). Teaching problem-solving skills. *American Mathematical Monthly, 87,* 794–805.

Schoenfeld, A. H. (1982). Measures of problem-solving performance and of problem-solving instruction. *Journal for Research in Mathematics Education, 13,* 31–49.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 484–494.

Simon, H., & Chase, W. (1973). Skill in chess.

American Scientist, 61, 394–403.

Slosson, R. L. (1981). *Slosson Intelligence Test.* New York: Slosson Educational Publications.

Slovic, P., Fischoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1–39.

Torrance, E. P. (1972). Can we teach children to think creatively? *Journal of Creative Behavior, 6,* 114–143.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition: Vol. 1* (pp. 205–232). New York: Academic Press.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Whimbey, A. (1975). *Intelligence can be taught.* New York: E. P. Dutton.

Received January 29, 1985
Revision received June 3, 1985 ■

Instructions to Authors

For further information on content, authors should refer to the June 1980 issue of this journal (Vol. 72, No. 3, p. 277). Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (3rd ed.). All manuscripts must include an abstract of 100–150 words typed on a separate sheet of paper. Typing instructions (all copy must be double-spaced) and instructions on preparing tables, figures, references, metrics, and abstracts appear in the *Manual.* Also, all manuscripts are subject to editing for sexist language.

APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more journals. APA policy also prohibits duplicate publication, that is, publication of a manuscript that has already been published in whole or in substantial part in another journal. Authors of manuscripts submitted to APA journals are expected to have available their raw data throughout the editorial review process and for at least 5 years after the date of publication.

Because the reviewers have agreed to participate in an anonymous reviewing system, authors submitting manuscripts are requested to include with each copy of the manuscript a cover sheet that shows the title of the manuscript, the authors' names and institutional affiliations, the date the manuscript is submitted, and footnotes identifying the authors or their affiliations. The first page of the manuscript should omit the authors' names and affiliations but should include the title of the manuscript and the date it is submitted. Every effort should be made by the authors to see that the manuscript itself contains no clues to their identities.

Authors should submit manuscripts in quadruplicate. All copies should be clear, readable, and on paper of good quality. A dot matrix or unusual typeface is acceptable only if it is clear and legible. Dittoed and mimeographed copies are not acceptable and will not be considered. Authors should keep a copy of the manuscript to guard against loss. Mail manuscripts to the Editor, Robert C. Calfee, School of Education, Stanford University, Stanford, California 94305.